

Qualitative Monitoring of Video Quality of Experience

Pablo Pérez¹, Jesús Gutiérrez², Jaime J. Ruiz¹, and Narciso García²

¹Alcatel-Lucent

María Tabau 9, 28050 Madrid, Spain
{pablo.perez, jaime.ruiz}@alcatel-lucent.com

²Grupo de Tratamiento de Imágenes

Universidad Politécnica de Madrid, 28040, Spain
{jgs, narciso}@gti.ssr.upm.es

Abstract—Real-time monitoring of multimedia Quality of Experience is a critical task for the providers of multimedia delivery services: from television broadcasters to IP content delivery networks or IPTV. For such scenarios, meaningful metrics are required which can generate useful information to the service providers that overcome the limitations of pure Quality of Service monitoring probes. However, most of objective multimedia quality estimators, aimed at modeling the Mean Opinion Score, are difficult to apply to massive quality monitoring. Thus we propose a lightweight and scalable monitoring architecture called Qualitative Experience Monitoring (QuEM), based on detecting identifiable impairment events such as the ones reported by the customers of those services. We also carried out a subjective assessment test to validate the approach and calibrate the metrics. Preliminary results of this test set support our approach.

Keywords— *Quality of Experience; monitoring; IPTV; broadcast; subjective assessment*

I. INTRODUCTION

Most subjective and objective multimedia quality metrics which are described in the literature are aimed at obtaining a numeric result (MOS, *Mean Opinion Score*), either globally for a specific multimedia sequence or continuously along time. The target of a MOS-based objective metric is estimating the results that subjective assessment tests would provide for the same test scenario. The usefulness of this approach has been proved for the comparison of different degradations from the same original source, e.g. to analyze the quality provided by a video coding algorithm. Therefore, MOS-based metrics have been mainly targeted to design and evaluate media encoders, as they are more accurate than simple Peak Signal to Noise Ratio (PSNR) analyses, and much more practical than subjective quality assessment sessions [1].

However, these methods present some limitations. For instance, in those metrics, the relationship between the measurement and the MOS value is not straightforward, and can be easily biased to fit specific scenarios. Thus, when objective metrics are tested against generic use cases, their results are modest. In general, *Full-Reference* metrics are the best-performing [2], while estimating MOS is usually more difficult with *Reduced-Reference* or *No-Reference* [3], which are the most suitable techniques for quality monitoring.

These limitations in multimedia-based approaches make that continuous quality monitoring is mostly done by network-based algorithms, focused on Quality of Service (QoS) parameters (loss rate, bitrate, latency, jitter...). Among them, the *de facto* standard in multimedia QoS measurement is the Media Delivery Index (MDI) [4].

Basically, it is just a composition of a loss rate and a jitter measurement, entailing an easy implementation and deployment in several network points. Besides, in aggregation (for long periods and a high number of users), multimedia quality is quite related to loss rate (either in the network or in the terminal due to jitter), which gives sense to the metric [5].

However MDI is quite a rough approach to the end user Quality of Experience (QoE), due to the following reasons:

- Not all the video losses have the same impact.
- Due to schemes which can protect differently parts of the video with different priority [6], similar loss rates can have different effects depending on which packets are lost.
- MDI values are expressed in network terms (“packet loss”) instead of experience terms (“problems in video visualization”). Hence it is difficult to extrapolate them to values which are meaningful either to the end user or to the service provider.

Therefore, there is a need to develop a metric which fills the gap between pure multimedia MOS-based metrics (usually complex and encoding oriented) and MDI (too coarse results), which also increases the meaningfulness of the values provided. The aim of this metric would not be directly estimating the MOS but, by keeping the advantages of QoS-based metrics such as MDI (lightweight to implement and easy to understand), overcoming their limitations.

In this paper we address this problem by designing a monitoring strategy which makes the best use of the information available at network level to provide significant quality data to the service provider, extending our previous work in this area [7]. First, we establish the requirements of such a system in Section II. In Section III, we design the system itself and define which metrics to use and how to integrate them. In Section IV, we present a specific subjective assessment methodology which is suitable to the scenario of continuous quality monitoring, and we provide some subjective results supporting the approach taken. Finally we point out some conclusions of the work.

II. PROBLEM STATEMENT

Service providers need systems to monitor the quality that is being perceived by the end users of their multimedia applications, such as in IPTV. This way, they can both make statistics about the offered quality, and address specific quality degradations for any user in the deployment. Those systems implement algorithms which somehow model the

QoE of the viewers and report their results according to them. These algorithms should be:

- Meaningful for the end users and the service providers. This means that QoE should be expressed in terms which are understood (and, as far as it is possible, agreed) by the final users and the providers of the services.
- Scalable so that it is possible to take measures in each of the user terminals in the network. These measures shall be easy to aggregate, as well as require small computing power so that they have no impact in the service.
- Capable to operate without requiring full reference from the original multimedia source, as it will not be available in the measurement point.

To cover the first two requirements, the system should be able to detect and measure typical impairments caused by transmission errors in packet-based networks. For example, macroblocking, screen freezes, choppy transitions (short freezes), audio errors, edge noise, or service outage [8]. The proposed approach is based on the detection of these kinds of events, their mapping to quality levels, and the validation of this mapping by using subjective assessments. According to the two last requirements, the impairment detectors used shall be lightweight and No-Reference (NR).

III. SYSTEM DESIGN

No-Reference quality estimators can be modeled as processes with three steps: measure, pooling, and map to quality [9]. The aim of our design is keeping a good traceability of quality among all these steps, so that all their results are meaningful for the user of the quality monitoring system (i.e., the service provider). This provides a guarantee that the service providers understand the system, and therefore, can actively recalibrate it to fit their own perceptions or requirements.

The basic idea of the proposed solution is using a qualitative approach to the measurement of multimedia QoE, which is done under the following assumptions:

- The whole end-to-end multimedia distribution system is designed in a way that it offers enough quality under normal operating conditions. This means that there is a target quality, implicit in the original design, which is offered by the service provider. This includes the multimedia encoding of average content.
- There are events which impair the quality of experience. The most typical is the loss of packets, which generates video macroblocking or audio glitches, among other consequences. Other events could be a change in video encoding quality, loss of lip synchronization, etc.

The aim of the metric is detecting these events, either directly (e.g. by measuring the video output of the system) or indirectly (by intermediate measures). Fig. 1 shows the block architecture of the design, which we will name *Qualitative Experience Monitoring* (QuEM).

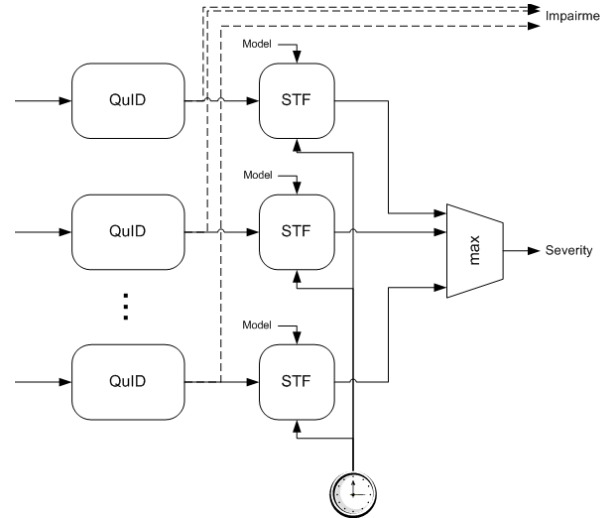


Figure 1. QuEM architecture design.

The basic building block of the solution is the Qualitative Impairment Detector (QuID). A QuID is an element which monitors the stream (either before or after decoding) in order to detect a single kind of impairment event.

A duration and a severity (which might be related with the duration) are assigned to each event. This makes possible the generation of statistics, which will be the QuEM results (mean, values, histograms...). Then, the operator of the solution can define a model for each kind of impairment which maps the values of the impairments to severity. This is the Severity Transfer Function (STF). STFs integrate the output of the QuIDs and provide a single severity value each T seconds. Thus, both pooling and map to quality processes are performed in this step.

The outputs of all the STFs are synchronized so that the aggregated severity is the maximum of all the severities involved [10]. A pooling window of 6 seconds is proposed, since it allows 10 measures per minute, which is easy to handle and aggregate by statistics processors. This output will be sent to higher-level aggregation systems, to generate statistic reports (e.g. histograms) of the occurrence and the severity of impairment events along time and over different clients.

It is worth noting that the STF could use even a MOS scale as severity measure. However, due to the traceability of the error measure, the mapping is done in terms of degradations which, by itself, are understood and described by end users. This way it is possible to compare the results of the QuEM with customer opinions.

A. Qualitative Impairment Detectors

The main events to detect are the kind of impairments which are valuable for the end users [8]. Table I shows these impairments, together with their main root causes and the QuID used to detect them. In order to establish the root causes, we have made two basic assumptions. First, the quality of the multimedia source is guaranteed and, therefore, the root cause must be either in the encoding/decoding process or in the network. If this were not true, then the

TABLE I. QUID FOR SEVERAL MULTIMEDIA IMPAIRMENTS

Impairment	Main Root Cause	QuID
Blockiness	Video Packet Loss	PLEP
Screen Freeze	Video Packet Loss	PLEP
Choppy Transitions	Video Packet Loss	PLEP
Distorted Audio	Audio Packet Loss	Audio Packet Loss
Edges shimmer	Video Compression	QP
Error stop	Outage	Outage

quality of the source should be monitored separately. Second, the audio coding quality is good enough and audio compression artifacts are negligible. Any good service design should fulfill this condition, as required audio bitrate is one order of magnitude lower than video bitrate for a similar quality.

The most severe effect of blockiness (and the one typically described by end users) is caused by errors in macroblock prediction due to a packet loss event resulting in the partial loss of a video frame. Packet losses can also cause the loss of complete frames, whose effect is the freeze of the video output for a short (choppy transitions) or long (screen freeze) period. All these events can be modeled with a Packet Loss Effect Prediction (PLEP) metric, such as the one proposed in [11], which also fulfills the design requirements of being a lightweight No-Reference tool.

With the assumption of audio coding quality being good enough, audio distortions can only be caused by packet losses. Since audio coding standards do not use temporal prediction as video standards do, the loss of audio packets produces a loss (silence) in the audio output signal of a similar length.

It is worth noting that video or audio packet loss effects can also be produced by jitter in the network. In such cases, the effective packet loss patterns may vary from the case where a real packet loss has been produced in the network. However, with a correct modeling of the client reception buffer in the monitoring system, this behavior can be easily detected and translated into effective packet losses. Then PLEP and audio packet loss models are applied.

Artifacts caused by coding quality can also be considered. The “edge noise” effect, as well as a part of blockiness effects, is caused by a strong quantization of the video signal. Quantization noise can be monitored in the bitstream, up to some point, by reading the QP value. These kinds of artifacts are the less reported by users [8], and some interesting studies of the relation between QP and perceived quality have been already published [12], so these impairments were not considered in our subjective test. Besides, video coding quality can be monitored on the encoder side using more accurate techniques (e.g. Full-Reference metrics), since this quality does not change along the transmission chain. In such case, this value can be distributed throughout the chain together with the video and used directly as output of the QuID element.

“Error stop”, described as the outage of the service (and blockage of the user terminal), can be caused either by a long interruption of video and audio stream or by hardware

failures of the network elements or client devices. The former case is an extreme version of packet loss events, which have been already considered. The latter is a kind of failure which is out of the scope of this work.

B. Mapping into Quality

The output of the QuIDs is a description of the impairment, such as “ x frames where $y\%$ of frame is impaired with macroblocking”. Once the impairment event has been correctly characterized, assigning a quality value is completely subjective. To help in this assignment, we have design subjective tests which try to find out the relationship between those identified impairments and their perceived quality. The results of these tests, which are presented in Section IV, make possible to define a STF for the QuIDs.

However, it is clear that any user of the QuEM architecture could modify the relation between QuID output and quality value just by modifying or redefining the STF. Due to the intrinsic limitations of generalizing the results of subjective assessments, this flexibility is a clear advantage of the QuEM architecture.

IV. SUBJECTIVE VALIDATION

To assure the validity of the proposed quality monitoring technique, subjective evaluation studies are essential, since people are the final users of the architecture. Furthermore, subjective tests are the basis of the process of mapping into quality the detected impairments, as they have to be in concordance with the results of the assessment tests.

From the point of view of multimedia service providers, the best way to know the QoE perceived by the end users of their systems would be having a direct report from their households while they are watching the content. Obviously, this is not feasible and it is necessary to turn to subjective tests to estimate the impact of possible distortions that may take place in the broadcasting system. However, these assessment tests should be designed respecting as far as possible real home viewing conditions, since this is the interesting scenario for service providers to evaluate the performance of their systems. This fact influences in the whole design of the subjective assessment methodology, from the viewing environment to the evaluation method, passing through the selection of test material. Therefore, the subjective assessment methodology should satisfy some requirements:

- Allow the evaluation of the different degradations of the QoE caused in the broadcasting chain.
- Keep as far as possible real viewing conditions of the end users of a broadcasting system (i.e. home environment).
- Make possible the comparison of the results of different experiments, being compatible with standard evaluation approaches.

Taking this into account, an assessment procedure was designed, and a subjective experiment was carried out to evaluate the impact of typical transmission errors (see Table I) on the quality perceived by the end users.

A. Proposed Methodology

As the main objective is to mimic home viewing conditions, the proposed methodology is based on standard single stimulus methods, such as those recommended by the ITU [13] and the Absolute Category Rating (ACR) [3]. In these cases, like when people watch video sequences at home, the observers do not have an explicit reference to compare with the content to evaluate.

To achieve a reliable evaluation of video quality from subjective tests, the material shown to the observers should be selected considering the properties of the sequences that the viewers usually watch at their households. This implies using test sequences that may be interesting for the subjects (e.g. movies, sports, etc.), so they will not be specifically focused in detecting impairments, as it happens in normal viewing conditions. Furthermore, in real situations long sequences are normally watched, thus, in the proposed method the length of the test material is a trade-off between having a sufficient duration to attract the attention of the observers and the limit of around 30 minutes for the total assessment session [14].

To make possible the evaluation of the impairments, the test videos are firstly divided into segments of 6 seconds, which is the length of the pooling window used in the proposed monitoring system. A diagram of the structure of the test sequences is shown in Fig. 2. The distortions considered for the tests are inserted in alternate segments (T_i). The non-distorted segments are labeled with a number (i) in the right-bottom corner of the image to indicate to the observers that they should rate the previous segment (distorted and unlabeled). A first segment labeled with a 0 is introduced to signal the start of the test and to give a ground coding quality of the video. In some cases, no degradations are inserted in T_i segments, thus they are evaluated as hidden references. Unlike ACR [3], this methodology allows a nearly continuous evaluation of the quality of the sequence without losing the continuity of the video, which simulates better real viewing situations.

The evaluation is done filling questionnaires according to the five-grade impairment scale proposed in [13]. The observers should write a cross in the square corresponding to the scale value and the segment rated, given by the number printed on the screen. During the non-distorted segments they have enough time to look to the questionnaires and rate without losing any distortion to evaluate.

B. Experiment Settings

The subjective assessment sessions were conducted in a room set according to the recommendations described in ITU-R BT.500-11 [13] for home environment tests. A 42" full-HD Panasonic television was used. In each session participated a maximum of 4 subjects, who were sat at a

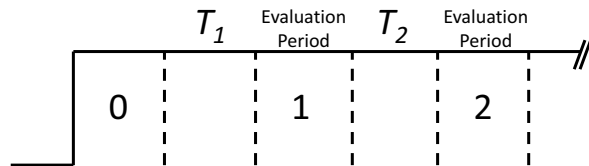


Figure 2. Diagram of the structure of the test sequences.

TABLE II. TEST SEQUENCES

ID	Sequence	Resolution	Framerate
SRC1	Movie	1920x1080p	23.976fps
SRC2	Documentary	720x576p	25fps
SRC3	Sports	720x576p	25fps

viewing distance of 3 times the height of the TV monitor.

A total of 19 observers took part in the experiment, all of them having normal visual acuity and color vision. No observers were rejected after the screening of the results following the indications of the recommendation [13].

Three video sources of five minutes were used in the test, whose main characteristics are shown in Table II. These sequences were firstly encoded in H.264/AVC and then encapsulated in RTP. One slice per frame was used and also a closed GOP of 24 frames, with a structure IBBBP, and setting all I-frames as Instant Decoder Refresh (IDR) pictures. The bitrate used for encoding the videos were 8Mbps for SRC1, and 4Mbps for SRC2 and SRC3.

At the beginning of each assessment session, the aim of the test and the evaluation methodology were explained to the subjects. In addition, a sample sequence was presented showing some examples of the distortions to evaluate.

C. Results

The results of subjective evaluation of the different types of impairments are described in the next subsections. The following figures show the MOS values for each impairment and each source sequence, and the global results considering the three of them.

These results will be used to propose a simple approach to be used in the STFs to map the qualitative outputs of the QuID to MOS values. Following the recommendation [13], a non-symmetrical logistic function is used. Its mathematical expression is given by (1), where a and b are the parameters of the function, D is the distortion output of the QuID, and u_{min} and u_{max} are the minimum and maximum subjective scores for the analyzed event.

$$MOS_p = u_{min} + \frac{u_{max} - u_{min}}{1 + (D/a)^{1/b}} \quad (1)$$

The specific distortion outputs (D) for each event are described in each of the following subsections. These outputs are obtained from the QuID which are based on the PLEP metric [11]. To obtain the parameters of the mapping function, the global MOS values (considering the three sequences) were used. The fitting process was carried out using Matlab and its command *fminsearch*. For validating the approach, the RMSE (normalized between 0 and 1, being 0 the indicator for best performance) was computed using the predicted MOS (MOS_p) and the MOS values for each one of the inserted degradations in the three sequences.

1) Blockiness

To analyze the effects of packet video losses on the perceived quality, the lost packets were chosen deterministically, instead of using random error patterns setting a loss rate. This is an appropriate way to study the

degradation caused by the lost packets depending on their position in the stream. The error concealment method considered in the decoder substitutes the lost macroblocks with the corresponding of the previous frame.

The impairments are generated by dropping one single packet of the I-frame of a GOP at different positions, causing macroblocking in the 25%, 50% and 100% of the frame. Due to the use of reference pictures in the decoding process, these errors propagate during one GOP (approximately 1s). Two more cases were analyzed to determine the influence of the duration of the effect. In these cases, a packet is lost in 3 and 5 contiguous GOPs affecting a 50% of the frame.

The results obtained in the subjective test for these impairments are shown in Fig. 3. As it can be expected, greater distorted regions cause a worse degradation on the perceived video quality. Also, the greater the duration of the impairment, the worse is its impact perceived. It is worth noting the high MOS values obtained for the SRC2, especially in the most severe cases, which is due to the type of content of the sequence. Usually, documentaries have scenes with low activity which cause that all the frames within one shot are very similar. In these cases, the error concealment method used for decoding makes nearly imperceptible the blockiness.

The output of the QuID in charge of the detection of these artifacts could be the portion of the frame affected by macroblocking (p) and the duration of this effect (t). Therefore, the mapping to MOS values could be done using the expression (1), and considering the total distortion parameter D as stated in (2), where α is a coefficient obtained as the other parameters of the mapping function:

$$D = \alpha \cdot p \cdot t \quad (2)$$

After applying the data fitting, a RMSE of 0.219 was obtained, which validate this simple mapping proposal for the STF of the architecture.

Probably, better results could be obtained using a more complex mapping function and a QuID based on a NR metric operating not only with the bitstream but also with the decoded frames. For example, in cases like those aforementioned concerning SRC2, the QuID output for the percentage of the frame affected would not be based on the bitstream but on the actual blockiness effects in the pictures. Thus, this will be focus of future work.

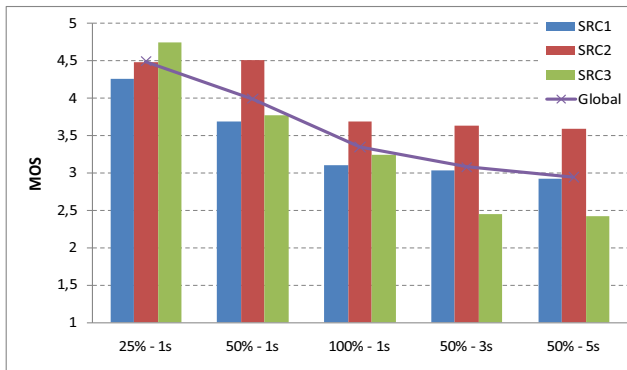


Figure 3. Effects of macroblocking.

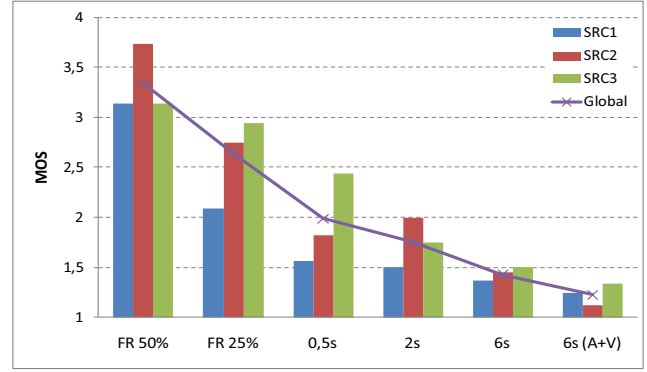


Figure 4. Effects of choppy transitions, video freeze, and outage.

2) Choppy Transition, Screen Freeze and Error Stop

The effects of discontinuities in the natural flow of the video were also under study in the test. Choppy transitions were simulated by reductions of the framerate to 50% and 25% of the original. Moreover, video freezes of 0.5s, 2s and 6s were inserted in the evaluated segments. Also a total outage event of audio and video during 6s was considered.

The subjective results obtained in the test are shown in Fig. 4. Discontinuities like changes in the framerate of long duration are less annoying than screen freezes, while total outage is the worst degradation in comparison with all the rest, since the observer miss completely the multimedia information. As expected, longer durations of video freeze are more annoying. Also in these situations the video content is an important factor, since depending on when the video freeze happens it could be more annoying. For instance, some of the freezes took place during a replay in SRC3, in which the camera is static being sometimes imperceptible. In SRC2 many shots of the video show static landscapes and this is the reason of the irregular results for video freezes.

These events could be easily detected with the PLEP metric, reporting the duration of the impairment. In addition, to model the effects of framerate drops, the distance between different frames when repetitions are detected could be used. In the case of video freeze this distance will be equivalent to its duration, but in the case of framerate drops, it will identify its severity. Therefore, a linear combination of these two factors will correspond to the total distortion parameter (D) to be used in the mapping described in (1), as shown in (3), where t is the duration and d is the distance between different frames:

$$D = \alpha \cdot t + \beta \cdot d \quad (3)$$

Fitting the data the coefficients α , β , a and b were obtained and comparing the outputs of the mapping with the MOS values of all the segments distorted with these impairments the RMSE was 0.175. Total outage has been not considered in the mapping, since it is as unacceptable as freezes of the same duration, and more exhaustive studies of the join effects of audio and video are needed.

3) Audio Distortions

In Fig. 5 the results concerning the effects of losing audio packets are shown. Different patterns and durations of audio

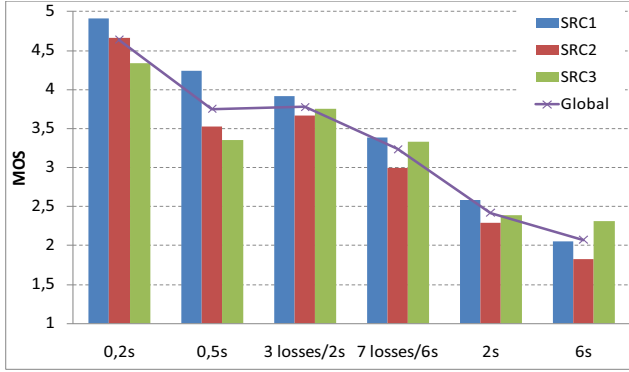


Figure 5. Effects of duration of audio losses.

losses were analyzed. The impact of losing a single audio packet (200ms of audio approx.) is compared with losing 0.5s, 2s and 6s of audio. In addition, two cases of various losses during a time period are considered: 3 losses of one single packet in 2s, and 7 losses of one single packet in 6s.

As it is expected, the longer the loss, the more annoying the impairment. Regarding the effects of having bursts of losses, having a single loss of t seconds is more annoying than having various minimum losses (only one audio packet) during t seconds. However, the effects of having various losses is comparable to having a loss of similar duration of the total of that various losses. For example, when 3 audio packet are lost, 3 silence periods of around 200ms are perceived (a total of 600ms), and the MOS results in this case is similar to that obtained for a single loss of 500ms.

Therefore, considering outputs of the QuID such as: " t seconds of audio were lost during T seconds", a simple approximation for modeling the effects for audio distortions in the STF block could be based on the total duration of audio losses. Using the mapping described by (1), with D equals to the duration of the impairment (t), the RMSE was 0.133 considering all the events of audio distortions.

V. CONCLUSIONS

Multimedia quality monitoring systems for content delivery services require some properties which are rarely shown by typical MOS-based media quality estimators. For that reason, service providers tend to use systems which implement network QoS. However, by using the proposed architecture QuEM, it is possible to enhance those metrics to provide QoE information without a significant increase of computation complexity.

QuEM model is based on metrics whose results can be described in terms which are significant to the end user and the service provider. They are also repeatable, and it is possible to simulate impairment events whose effects are known *a priori*. Each measure subsystem, or QuID, is independent from the others, thus making it possible to improve them individually without affecting other parts of the architecture. And the result of this QuID can be mapped to quality values using a STF which can be modified or redefined by the QuEM user. For all these reasons, QuEM is a rather powerful architecture, especially suited for real-time monitoring of complex multimedia services.

A subjective assessment methodology has been defined to evaluate the performance of this approach. It is based on the idea of carrying out the quality evaluation in the most realistic way compared to home environment viewing conditions. Preliminary results of this test support the design of the architecture and validate the provided qualitative results. Also, they show that the types of errors that have been under study are reasonably significant for the users.

Future work will focus on the development of efficient QuID and STF elements, and on validating the error pooling and mapping architecture with new subjective results.

ACKNOWLEDGMENT

This work has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government under project TEC2010-20412 (Enhanced 3DTV), and by the Ministerio de Industria under project ITEA2-JEDI.

REFERENCES

- [1] J. You, U. Reiter, M.M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio-visual services: A survey", *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 482-501, Aug. 2010.
- [2] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment, Phase II", 2003.
- [3] VQEG, "Validation of reduced-reference and no-reference objective models for standard definition television, Phase I", 2009.
- [4] J. Welch, and J. Clark, "A proposed Media Delivery Index (MDI)," *Internet Engineering Task Force (IETF) RFC 4445*, 2006.
- [5] A.R. Reibman, V.A. Vaishampayan, and Y. Sermadevi, "Quality monitoring of video over a packet network," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 327-334, Apr. 2004.
- [6] D.C. Robinson and A. Villegas, "Intelligent wrapping of video content to lighten downstream processing of video streams," *European Patent EP2071850*, 2009.
- [7] P. Pérez and N. García, "Lightweight multimedia packet prioritization model for unequal error protection," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 1, pp. 132-138, Feb. 2011.
- [8] G.W. Cermak, "Consumer opinions about frequency of artifacts in digital video," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 336-343, Apr. 2009.
- [9] S.S. Hemami, and A.R. Reibman, "No-reference image and video quality estimation: Applications and human-motivated design," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 469-481, Aug. 2010.
- [10] S. Jumisko-Pyykkö, V. Kumar, and J. Korhonen, "Unacceptability of instantaneous errors in mobile television: From annoying audio to video," *Proc. Conf. on Human-computer interaction with mobile devices and services*, pp. 1-8, Sep. 2006.
- [11] P. Pérez, J. Macías, J.J. Ruiz, and N. García, "Effect of packet loss in video quality of experience", *Bell Labs Technical Journal*, vol. 16, no. 1, pp. 91-104, Jun. 2011.
- [12] M. Barkowsky, M. Pinson, R. Pèpion, and P. Le Callet, "Analysis of freely available subjective dataset for HDTV including coding and transmission distortions," *Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Jan. 2010.
- [13] ITU, "Methodology for the subjective assessment of the quality of television pictures", *Recommendation ITU-R BT. 500-11*, 2002.
- [14] N. Staelens et al., "Assessing quality of experience of IPTV and video on demand services in real-life environments", *IEEE Transactions on Broadcasting*, vol. 56, no. 4, pp. 458-466, Dec. 2010.